Software Tools for Sharing and Integrating GIS Data

Project Description: Created for the Washington State Transportation Framework

(WA-Trans)

Last Updated: June 15, 2009

Washington State Department of Transportation

[THIS PAGE INTENTIONALLY LEFT BLANK]

TABLE OF CONTENTS

SUMMARY 4

BACKGROUND 5

CONSORTIUM FOCUS AREAS (ONE-ROAD PILOTS) 6

WA-TRANS SOFTWARE AND PROCESSES (ONE-ROAD PILOTS) 7

DATA PROVIDER INTERNET PORTAL (ONE-ROAD PILOT PHASE I) 7

TRANSFORMATION (PUGET SOUND PILOT PHASE I AND ONE-ROAD PILOT PHASE I) 8

CHANGE DETECTION (ONE-ROAD PHASE I – REQUIREMENTS AND FEASIBILITY, ONE-ROAD PILOT PHASE II – IMPLEMENTATION) 8

DATA INTEGRATION (ONE-ROAD PILOT PHASE I – REQUIREMENTS & FEASIBILITY, ONE-ROAD PILOT PHASE II – IMPLEMENTATION) 9

QUALITY ASSURANCE AND QUALITY CONTROL (ONE-ROAD PILOT PHASE II) 10

SECURITY (ONE-ROAD PILOT PHASE I ENHANCED IN ONE-ROAD PILOT PHASE II) 10

DATA USER INTERNET PORTAL (ONE-ROAD PILOT PHASE I) 10

LOCATION REFERENCING INTEGRATION 11

THE NEED FOR A CONSORTIUM 11

CONSORTIUM BENCHMARKS 12

CONSORTIUM MANAGEMENT 12

CONTACTS 13

Summary

The Washington Department of Transportation (WSDOT) in partnership with the six other state departments of transportation are involved in a consortium of public and private entities. The consortiums purpose is the development of computer based tools that facilitates geo-spatial transportation data sharing and integration providing a multitude of business benefits. The Geo-spatial Integration and Sharing Data Consortium (GISDC) is funded via a Transportation Pooled Fund (TPF) arrangement and managed by WSDOT.

The consortium is actively seeking the participation of other state departments of transportation who will benefit from the results of this project.

The goal of the consortium is to research, develop and implement a variety of software tools and procedures for sharing and integrating geo-spatial transportation data. There is a need to share data among state, county, city, and other jurisdictions in order to significantly reduce the time, effort, and expense of transportation projects in which integrated multi-jurisdictional and/or multi-modal data is required.
The primary goals of the project are:

- Design and implementation of a database within the state as a central repository of core multi-jurisdictional GIS and tabular transportation and location data,
- Transformation of GIS and tabular transportation data from disparate spatial data formats and schemas into a centralized database,
- A web portal designed to securely receive data from data providers and to facilitate regular maintenance updates,
- Change Detection to facilitate provider maintenance updates,
- Integration of data from disparate data sources into a seamless whole,
- QA/QC processes to monitor data quality, security, data entry and retrieval processes,
- A web portal to provide data users with integrated multi-jurisdictional, multi-modal state wide GIS transportation and location data,
- Documentation of a set of processes necessary to support data sharing from a variety of sources, e.g. data sharing agreements, agreement points,
- Documentation to support all technical aspects of the project,
- Research and implementation of data schema and processes to support Linear Referencing Integration.

Because it is necessary to include as many agencies and jurisdictions as possible, the software tools and processes need to work as a dynamic system that can accommodate numerous data structures in a wide range of hardware and software environments.

The flexibility required to achieve the project goals will facilitate this set of tools being used, with minor modifications, in virtually any state transportation department. This is especially true if that state is a participant in the shared pool. The more participants and resources involved, the easier it is to develop the most dynamic and flexible tools.

Data to be managed with these software tools and processes includes:

- Roads: State and Local roads, ramps, rest areas, location, number of lanes, federal functional classification, ADT, speed limits, address ranges, zip codes, local road identifier, route number, road name(s), location along roadway (milepost and/or GPS), and their geographic representation,
- Railroads: location, type of track (mainline, siding, etc.), train stations location, classification, line identifier, type of crossing, and their geographic representation,
- Ferries: route location, terminal location, route name, federal functional class, staging areas, route length, international or domestic route, average sailing duration, etc.,
- Aviation: Air Port identifier, surface type. Instrument landing approach, arc code, runway width, use, elevation, FAA Classification, Airport name, terminal location, etc.,
- Non-motorized: includes bikes, foot, horses, etc. Includes location, name, type of usage, etc.,
- Ports: location, routes for water transportation (particularly river and Puget Sound),
- Various jurisdictions including, but not limited to: cities, counties, tribal nations, various state, local and federal agencies,
- Other data as yet not defined.

The initial timeline for this development is projected to be three to five years. Annual consortium meetings, quarterly conference calls and electronic communication are being used to coordinate project activities. The specific product specifications and the scope of the project depend on annual funding available.

Background

For state transportation agencies location is an integral part of most data collected and utilized. Data with a location referencing element (e.g. address, route/milepost, GPS coordinate) can be used with a Geographic Information System (GIS) and placed on the roadway or other parts of the transportation system. Much of the useable data about transportation is not maintained at the state DOT level, but at the county, city or other local level or with other agencies. For planning, project scoping, environmental management, emergency management and other integral DOT functions it is often necessary to collect and combine this data. Current circumstances make it difficult, time consuming, and expensive to do so. Many of these difficulties could be mitigated by the creation of a complete statewide transportation network and associated location referencing systems.

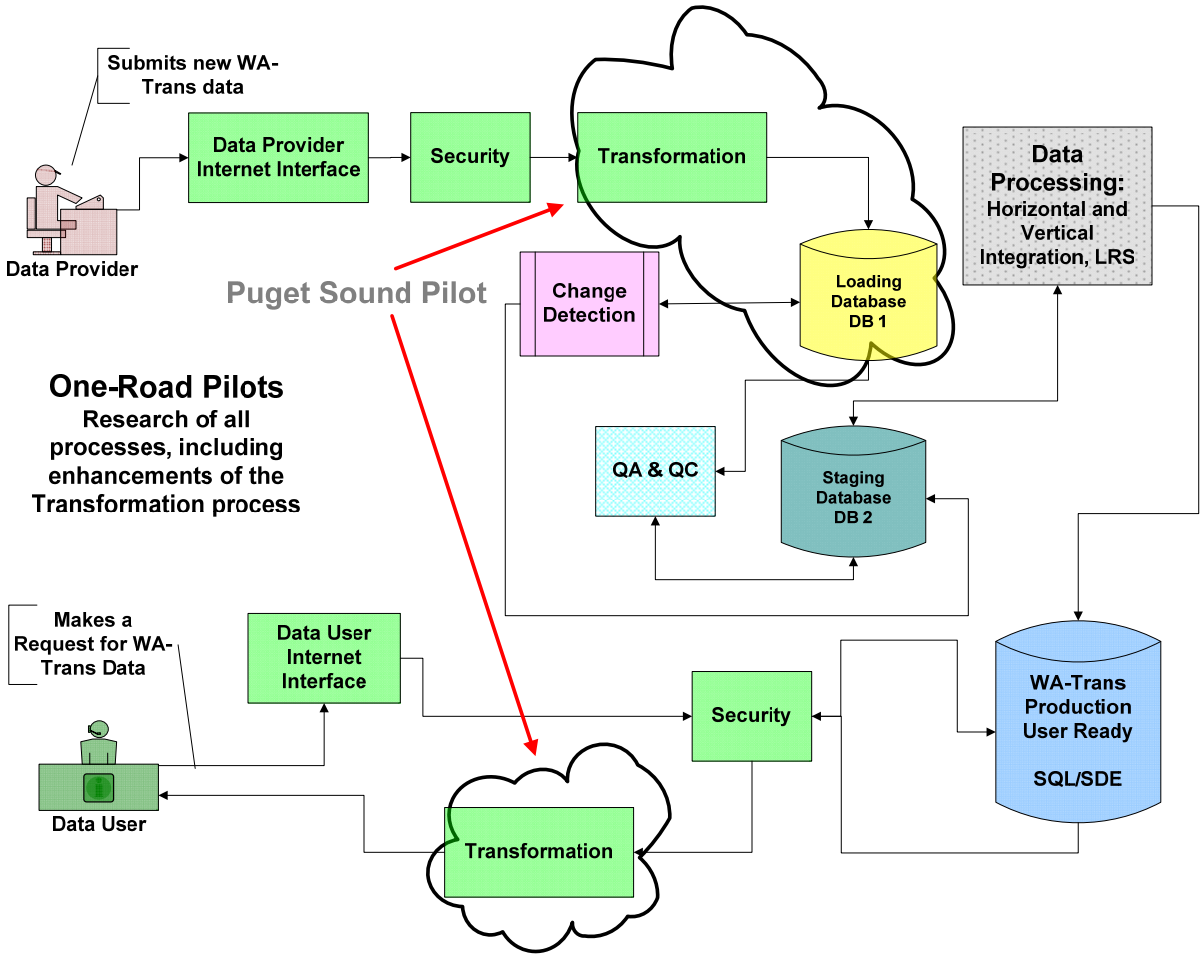
The Oregon Department of Transportation (ODOT) and the Washington State Department of Transportation (WSDOT) have been working on projects to collect and integrate local and state agency data statewide, for GIS, known as a “Transportation Framework” for some time. The Ohio Department of Transportation (ODOT) and California Department of Transportation (CalTrans) are also well into transportation network projects. In Washington the project is called “WA-Trans”, for Washington State Transportation Framework for GIS.

More states became involved as interest in this research increased. Current members of the GISDC are:

- Oregon Department of Transportation (ODOT),
- Ohio Department of Transportation (ODOT),
- Tennessee Department of Transportation (TDOT),
- Nebraska Department of Roads (NDR),
- Idaho Transportation Department (ITD) and the
- California Department of Transportation (CalTrans)

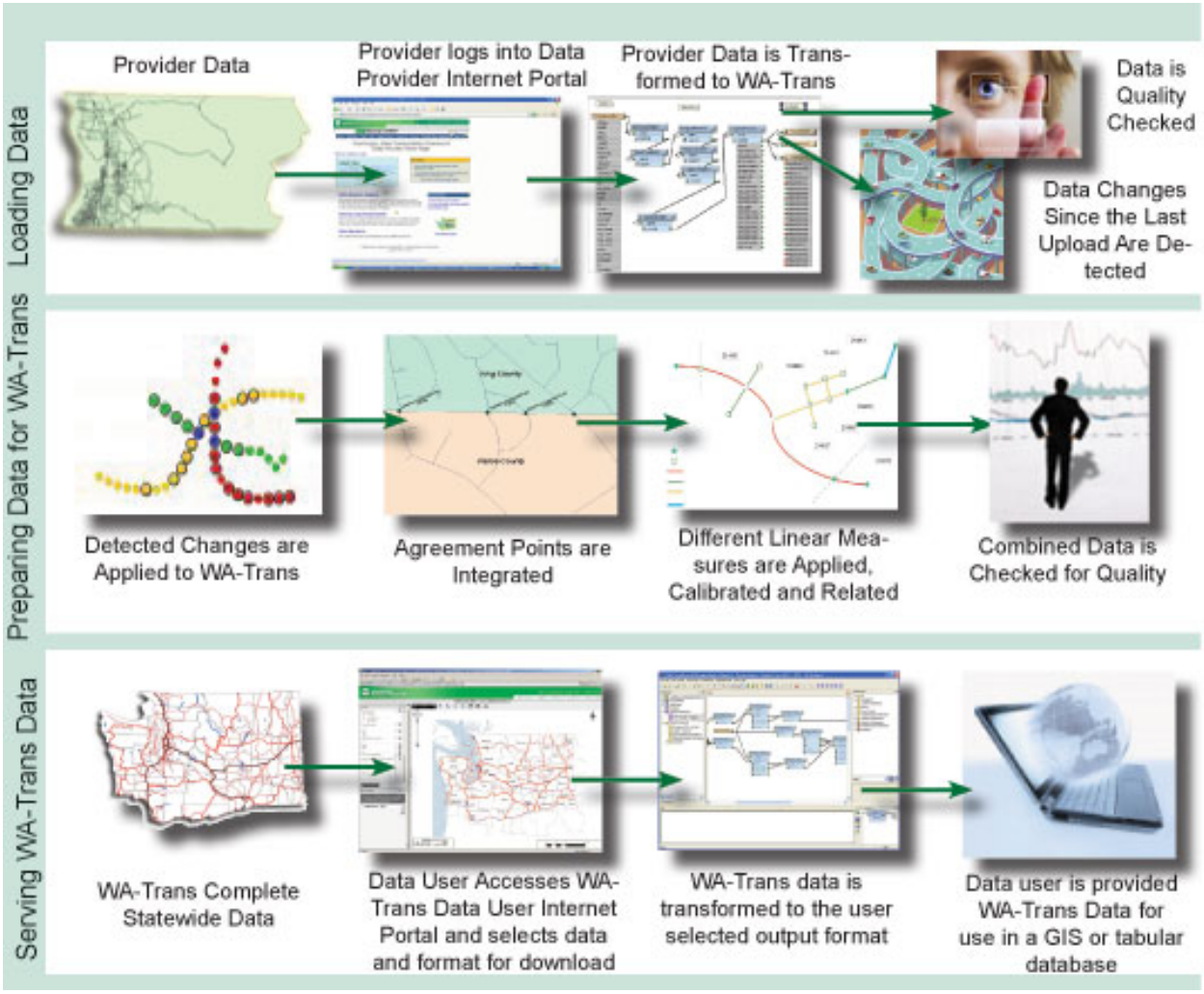
Consortium Focus Areas (One-Road Pilots)

The diagram below depicts and basic architecture and the flow of data from a data provider through the databases and back out to a user.



WA-Trans Software and Processes (One-Road Pilots)

Software processes are needed to move the disparate data submitted into some standardized form and then serve that data out to a user. The original meaning of the data cannot be altered during this process. Each of the consortium focus areas are being researched to this end. Of the shelf software will be considered when possible. The diagram below indicates processes necessary to receive provider data, manage that data and then serve that data to a user.



A description follows of each proposed tool, and how it may interact with the other tools. The consortium is currently involved in the One-Road Pilot Phase I and Phase II.

Data Provider Internet Portal (One-Road Pilot Phase I)

Data providers will access an Internet interface, which will allow them to upload data to WA-Trans. Transformation processes will be automated, after initial creation. This will facilitate a quick and user friendly method to provide data to WA-Trans. A provider will be able to upload and update the framework data as their local data changes. Updates will probably occur on a regular basis, based upon arrangements between data providers and WA-Trans. The Data Provider Interface will also allow a provider to modify the transformation parameters if there are changes within their systems. For example, if a provider adds new attribution to a GIS file being

submitted they will be able to alter the transformation parameters using functions available in the Provider Interface to reflect the newly added attribution. Security procedures will be in place to prevent unauthorized access to the data provider interface.

Status: *The requirements are complete. A prototype interface is being developed. A very rudimentary interface testing the concept of data flow and server upload processes is complete. The prototype interface is anticipated to be completed September, 2008.*

Transformation (Puget Sound Pilot Phase I and One-Road Pilot Phase I)

A major part of the Transportation Pooled Fund Project involves the development of a dynamic, flexible, bi-directional translator that will transform data from standard GIS file types (e.g. Shape, Coverage, CAD) to a common schema. The transformation process will be able to convert GIS vector data with attribution and location referencing from one format into another. The transformation processes will be designed to be reused and to update and maintain the data.

In the Puget Sound Pilot transformation processes were researched with GIS ETL software products. The One-Road Pilot has taken the lessons learned in the Puget Sound Pilot and refined the transformation processes with several jurisdictions successfully. The transformation will also need to check the data for simple constraints to ensure it meets basic requirements as defined by the standards agreed to by the parties involved. The One-Road Pilot Phase I is currently integrating the transformation processes into Data Provider and Data User Web Portals.

Status: *Translations have been successfully completed.*

- *Translation from data provider's schema into the statewide database has been implemented for multiple providers.*
- *Translators out of the statewide database back to data providers in their own schema or into a standard schema have been implemented.*
- *Data has also been successfully translated to the National Map.*
- *A translator template has been created to facilitate creation of translators for new providers. This template has reduced creation time from the initial research of time of 200 hours per provider to between 20- 40 hours per provider.*

Sever-level transformation processes have been implemented and are ready to work with the Data User and Data Provider portals.

Change Detection (One-Road Phase I – Requirements and Feasibility, One-Road Pilot Phase II – Implementation)

A critical process for data maintenance is change detection. The data goals of WA-Trans include the ability to update transportation data on a regular cycle in order to maintain a current integrated transportation dataset. The ability to view history as well as current data is important. As providers upload a more recent dataset there is a need to only update the changes to the statewide database.

The Change detection is divided into two distinct parts:

- Detection of changes,
- Implementation of the detected changes.

Status: *Prototype change detection processes have been developed which support data being maintained from a variety of data providers. Change detection reports are being reviewed by providers. Providers have found these reports very helpful. Some work still remains on these items. Work is also underway to implement processes for applying changes that have been detected to the statewide database thus supporting long-term maintenance using shared data. It is expected these utilities and processes to be completed by July, 2009.*

Data Integration (One-Road Pilot Phase I – Requirements & Feasibility, One-Road Pilot Phase II – Implementation)

The data that will be provided from the various agencies will be linear, point and tabular data that represents roads and other transportation modes, with attributes that describe characteristics of the roads such as street/highway names, federal functional class, pavement type, etc. The provided data will also include location referencing information such as addresses, route names, and mileposts. Location referencing will allow other data, such as collision data, to be assigned to specific places along the transportation segment.

Because data comes from various sources and each source collects and stores data differently, it may not "match" at jurisdictional boundaries. When displayed on a map a road in one county may appear to just end when the same road in another county may appear to just begin. One road may be disjointed, and doesn't connect at all where it crosses a jurisdictional boundary, when in reality it is a single continuous roadway. Even within a specific jurisdiction there may be multiple providers of road data causing similar problems. Fixing this problem is referred to as "edge matching" or horizontal integration. An Agreement Point Process, developed in the Puget Sound Pilot and fine tuned in the One-Road Pilot Phase I, is an integral part of the "Edge Matching' in WA-Trans.

There are other types of mismatch. Making sure that the most accurate and complete database attribution is correctly linked and connected across various transportation modes (highways, rail, ports, airports) or to other "levels" (the power grid, water lines, cadastral information, etc) is referred to as vertical integration. Tools built to facilitate both horizontal and vertical integration are useful not only for a transportation framework, but for any linear based GIS data and related attributes. Early work on integration processes has begun.

Status: *Several integration processes are being researched with preliminary results expected this summer. Some processes being worked on now include:*

- *Comparison of two or more datasets from the same jurisdiction.*
- *Coincident features near jurisdictional boundaries and application of attribution from one feature to another.*
- *Comparison of State Route data within a provider's jurisdiction and implementation of State Route data within that jurisdiction.*
- *Re-Segmentation and application of attribution accordingly.*
- *Agreement points and Disagreement Points and implementation of those points in the transportation framework.*
- *Application of multiple descriptions for any feature and serving this out to providers.*

Manual processes to support some of these processes are near completion. Automated utilities development is just beginning.

Quality Assurance and Quality Control (One-Road Pilot Phase II)

Quality assurance and quality control (QA/QC) processes and tools establish and enforce data consistency and data accuracy. This is especially critical in an environment where data is being integrated from multiple sources and distributed to multiple users. Much of the QA/QC function can be automated in the following categories:

- Topological – checks regarding connectivity of the line work at intersections, overpasses and bridges represented as separate features, arcs meeting at jurisdictional boundaries, etc.
- Scale/Spatial – Does the location accuracy meet the planned business use of the data; does the "aesthetic" representation of the transportation feature meet the business requirements?
- Attribute – Are the minimum required fields included, are the field descriptions met, how many of the attributes are populated, are the attribute values valid?
- Metadata – Concerns regarding metadata include: has the required metadata been provided, is it complete, does it conform to established metadata standards; does the metadata match the layer?

Status: *Requirements are being identified. A strategy for when to perform specific types of quality control has been identified. Requirements for each are being developed. At that point opportunities for automation will be identified. In areas where automation is not useful or appropriate, "manual" processes will be established and documented.*

Security (One-Road Pilot Phase I enhanced in One-Road Pilot Phase II)

Data security is required for both Data Provider and Data User Portals. Data uploads and modifications may only be done by authorized providers. In addition, data entry and modification must be done through a secure system. While it is anticipated that transportation framework data is generally publicly available, some business situations (such as statewide E-911 dispatch) require that private data be used. Thus access to sensitive data must be secured against anyone except those authorized to use it. For uses other than transportation framework it is important that security be available.

Status: *Security has been implemented in the Data User and Data Provider Portals prototypes.*

Data User Internet Portal (One-Road Pilot Phase I)

Data users will be provided with an Internet interface allowing access to the fully integrated statewide transportation framework data. Users will be able to view and download data in their own GIS file types and formats using the transformation process within the Data User Portal. A user will be able to:

- Select the geographic region for the data they wish to view/download,
- View the metadata for that selection. Metadata is information about data such as the time the data was collected, the spatial accuracy of the data, the projection and coordinate systems of the data etc. NOTE: Metadata from the original providers and for the current data will be maintained.
- View the geographic data they have selected.
- Download the data in a user selected format.

A disclaimer will be provided regarding the limitation of the data. Transformation processes will be available for formatting the data and projecting it as needed by the data user.

NOTE: The Data User Portal is a "read only" interface that allows viewing and downloading data. Once they have downloaded the data they can then incorporate that back with their own GIS transportation data and use it accordingly.

Status: *The Data User Portal prototype is complete. There are several requirements that can't be implemented on a prototype implementation due to the test infrastructure, but the basic requirements are implemented. The ability to obtain specific data products is currently being researched.*

Location Referencing Integration

When building GIS for transportation infrastructure, a major business need is to be able to locate things along the infrastructure network. This is achieved through location referencing. Washington State employs multiple forms of location referencing:

- States and counties can use a form of route/milepost points for location referencing,
- Counties and cities may use addresses,
- Cities may also use distance from intersection
- Many may use GPS for various purposes.

In order to accurately place data events and features on a roadway, some form of linear referencing is necessary. In order to relate data event locations to each other, it is necessary to be able to relate the linear reference of each to the other. This will facilitate geocoding across the state and locating things by a variety of methods, meeting a variety of business needs. Software tools that support location reference integration will be critical when bringing outside data of any sort into any GIS system and then trying to use that data to locate features along a transportation network.

Although local location referencing will be provided with data, that location referencing will not always be consistent between data providers and across boundaries. In order to have a single location referencing system for the whole state, a consistent system must be applied to the data during integration.

Status: *We have added multiple LRS descriptions for a feature. However, utilities to support maintenance of this have not been designed. There are several issues to be resolved initialy, specifically the determination of a core set of segments. This is currently be researched in the conflation/integration research tasks and we expect to have a process developed by the late summer of 2009. Once the segmentation is resolved we will be able to apply multiple LRS data effectively.*

The Need for a Consortium

In order to complete the full suite of tools, processes, and documentation, a consortium of supporting members needs to provide consistent, continuing support, feed back and funds for a period of three to four years. It is hoped that members will also provide creative input, possibly time for testing, and other processes that will move the project forward. The consortium will provide implementation opportunities for tools developed during this effort and the documented

processes that establish the foundation for an integrated, maintainable statewide transportation network for each interested member.

Consortium Benchmarks

The GISDC is designed to be action oriented with the primary focus being the development and implementation of tools that facilitate sharing and integration of geo-spatial transportation data and testing of those tools with a variety of data formats and sources from a variety of geographic locations. It is important that complete documentation of all tools be provided including through conferences and refereed publications. Thus, papers will be developed as appropriate and submitted to organizations such as the Transportation Research Board (TRB), GIS-T and URISA for review, presentation, and publication. It is hoped that the Consortium members will co-author these papers. Consortium members will directly benefit from the software tools and processes that can be designed to meet their specific needs.

Consortium Management

The pooled fund lead states are Washington and Oregon. Funding for the Consortium is arranged through WSDOT through the Transportation Pooled Fund. Each consortium member has designated one individual to serve on the GISDC Advisory Team. The team guides the work of the GISDC. Meetings are held annually and electronic communications in addition to periodic conference call meetings serve to bind the team together between meetings.
Some of the tasks of the GISDC are to:

- Gain knowledge of existing software to determine which requirements can be met with off-the-shelf products, either as is or customized, and which requirements will require software developed from scratch, or COTS customization.
- Establish minimum requirements for each tool.
- Recommend the preferred way of communicating between the GISDC members and the research team.
- Establish periodic meetings and workshops. Day to day operations will be guided by interactions with the lead states of the Consortium.
- If consortium members choose to actively participate in the Pilots the GIDSC will provide guidance and communication processes. Active participation in Pilots by all consortium members is encouraged.

Contacts

For further information contact:

Tami Griffin (Washington)
griffit@wsdot.wa.gov
360.596.8926

Michael Leierer (Washington)
LeiereM@wsdot.wa.gov
360.596.8927

Dennis Scofield (Oregon)
Dennis.J.SCOFIELD@odot.state.or.us
503.986.3156

Chad Brady (Oregon)
chad.w.brady@odot.state.or.us
503.986.3164

Dave Blackstone (Ohio)
Dave.Blackstone@dot.state.oh.us
614.466.2594

Braun, Rose (Nebraska)
rbraun@dor.state.ne.us
402.479.3696

Kim McDonough (Tennessee)
kim.mcdonough@tn.gov
615.741.4037

Harold Feinburg (California)
harold_feinberg@dot.ca.gov
916.654.6585

Brian Emmen (Idaho)
brian.emmen@itd.idaho.gov
208.332.7889